22.8 High-Speed and Low-Energy Capacitively-Driven On-Chip Wires

R. Ho, T. Ono, F. Liu, R. Hopkins, A. Chow, J. Schauer, R. Drost

Sun Microsystems, Menlo Park, CA

Under VLSI scaling, on-chip wires present increasing latency and energy problems. As wire cross-sections shrink, repeated global wire delays will eclipse transistor delays [1] and may limit system performance. Also, increasing wire density with constant wire capacitance and supply voltage implies growing energy costs for global communication. Microstrip on-chip transmission lines are a possible solution [2]. They offer high bandwidth, but less bandwidth per area than densely packed repeated wires. They offer near-speed-of-light latency, but only at low bandwidth, as data serialization adds delay. Other solutions offer moderate repeater improvements in delay or power [3,4] or cut voltage swing for 10× energy savings at a 30% delay penalty [5]. Low-swing systems often need a second power supply, costing extra voltage-regulator modules at μΩ impedances and fractured chip/package power planes.

We reduce both latency and energy for on-chip wires by driving them capacitively. The capacitor pre-emphasizes transitions to decrease wire delay and reduces the driver's load to dramatically cut driver size and energy [6]. It also lowers the wire's voltage swing without a second power supply. Analysis shows that a capacitively driven wire is slightly faster and much lower power than a repeated 10mm global wire. Measured results in 0.18μm CMOS show 10.5× energy savings at a 50mV swing compared to full-swing repeated wires, with a 3× gain in wire bandwidth.

Driving a wire through a capacitor reduces signal swing through a capacitive divider. Figure 22.8.1 shows a reduced swing V_s at the wire end. Node B overshoots and later settles to V_s. Approximating the step response at node C with a single pole leads to a time constant

$$\tau \approx R_d(C_1 + C_c) + R_w(C_2C_3)/(C_2 + C_3) \approx R_d(C_1 + C_c) + R_wC_w/4$$

This is faster than an inverter driving a wire because the driver sees only C_c, not C_{wire}, and because charge redistribution between C_2 and C_3 effectively cuts the wire delay in half. While this simplistic model offers quick delay estimates and circuit intuition, it is pessimistic because the π-wire model limits the pre-emphasis spike at node B to $2V_s$. A distributed wire model more accurately shows a larger pre-emphasis spike at node B, which, in turn, charges the wire faster (Fig. 22.8.1). The speed-up of capacitively driven wires can also be seen from the wire's frequency response, as C_c introduces a pole-zero pair that boosts the wire's 3dB bandwidth. Figure 22.8.2 shows the simulated AC response of a 14mm long minimum-width wire where the capacitive drive reduces the gain but extends the bandwidth by over 3×.

We make C_c from wire sidewall capacitances to track C_{wire} under process variations. We use a pitchfork, shown in Fig. 22.8.3, with multiple tines and possibly multiple metal layers. For a 50mV swing in a 1.8V system, a 1cm wire with three driver tines and four wire tines requires a 50μm-long pitchfork. In practice, pitchforks are shorter because they use minimum wire-to-wire spacing, unlike the downstream long wire. The multiple wire tracks in pitchforks represent a relatively small overhead. Because the wire bandwidth extension reduces the number of repeater stages, we amortize the pitchfork cost over a longer wire.

Signals are sent differentially on twisted wires to enable sub-50mV swings; the twisting obviates shield wires by canceling wire-to-wire coupling noise (Fig. 22.8.3). The receiver is a StrongArm latch [7], sized for a 3.5σ offset under 30mV. Because C_c blocks DC, the wires and receiver inputs require a voltage bias. Our target application assumes DC-balanced data traffic, allowing us to bias receiver inputs using a leaky, off PMOS transistor with a time constant about 1ms. As long as data is DC-balanced at that timescale, the inputs will not drift far from V_{DD}. Systems without DC-balanced data can periodically precharge the channel. Multiple drivers, each with their own capacitor C_c, can drive the same long wire and thus sum their displacement currents. This enables a simple multi-tap FIR filter for driver-side pre-equalization (Fig. 22.8.4) [8].

Figure 22.8.5 compares repeaters to capacitively driven wires for a 10mm link in a 0.18μm process. These simulations varied driver size for the repeated links and varied coupling capacitance for the low-swing link. Repeated wires, each as wide as a pair of the low-swing wires, include a latch at the end; similarly, the low-swing link adds a sense-amp/latch. Both include an extra gate delay for clocking margin. The capacitively coupled system is slightly faster and 10× lower power than a repeated wire.

We built a 6-layer Al, 0.18μm chip with both capacitively driven and repeated 5b buses. Buses are 5mm to 8mm long and run at 1GHz, with voltage swings from 50 to 200mV. Ring oscillators underneath the buses inject random digital switching noise. An unrepeated, minimum-width 14mm wire provides a slow bus for bandwidth measurements, as well as a test site for a two-tap FIR driver. Figure 22.8.7 shows a particularly large 200mV pitchfork next to an 8mm bus wrapped back and forth eight times; 50mV swing pitchforks are 4× smaller. The target system strobes both drivers and receivers with the global clock, but our prototype uses separate clocks for testing.

Bus data comes from either a custom 5kb SRAM or from per-bit on-chip 31-stage PRBS generators. The SRAM allows testing of data patterns with worst-case noise; those experiments show the wire twisting effectively cancels wire-to-wire coupled noise. Every bus has at least a 50% UI window where the (test-time-limited) BER < 10^{-11}. The repeated buses use single-ended wires of the same pitch as a pair of the differential low-swing wires, and consume 1.1pJ/bit/mm. Although simulations show the low-swing buses to be faster, our test setup does not accurately measure latency due to unknown clock insertion delays. Low-swing buses targeting a 200mV swing (V_{DD}/9) consume 0.28pJ/bit/mm, a savings of 3.8× compared to the maximum savings of 4.5×; the difference comes from the sense-amp receiver and driver-side wire parasitics. At a 50mV swing, the low-swing bus saves 10.5× over repeated wires. At this voltage swing, the receiver, whose energy is not swing dependent, limits the energy savings.

Figure 22.8.6 shows waveforms on the slow 14mm wire experiment, probed using on-chip samplers [9]. While this wire is RC-limited to 55MHz, capacitive pre-emphasis enables 200MHz data with an adequate eye opening. The top graph shows the effects of a two-tap 5ns FIR filter. Simulations show that a tap delay of 1.5ns would have doubled the wire bandwidth; however, an unadjustable tap delay of 5ns gives only a 20% benefit and a minor increase in eye opening.

Driving on-chip wires with capacitors improves both latency and energy. Silicon results show 3× more bandwidth and 10.5× less energy at 50mV swing over full-swing CMOS repeaters.

Acknowledgements:
We acknowledge support from DARPA contract NBCH30390002.

References:
[1] R. Ho, K. Mai, and M. Horowitz, "The Future of Wires," *Proc. IEEE*, vol. 89, no. 4, pp. 490-504, Apr., 2001.
[2] A. Jose, G. Patounakis, and K. Shepard, "Pulsed Current-Mode Signaling for Nearly Speed-of-Light Intrachip Communication," *IEEE J. Solid State Circuits*, vol. 41, no. 4, pp. 772-780, Apr., 2006.
[3] H. Kaul, and D. Sylvester, "Low-Power On-Chip Communication Based on Transition-Aware Global Signaling," *IEEE Trans. VLSI Systems*, vol. 12, no. 5, pp. 464-476, May, 2004.
[4] A. Nalamalpu, S. Srinivasan, and W. Burleson, "Boosters for Driving Long Onchip Interconnects," *IEEE Trans. CAD IC Systems*, vol. 21, no. 1, pp. 50-62, Jan., 2002.
[5] R. Ho, K. Mai, and M. Horowitz, "Efficient On-chip Global Interconnects," *IEEE Symp. VLSI Circuits*, pp. 271-274, Jun., 2003.
[6] R. Drost, R. Ho, T. Ono, "Method and Apparatus for Driving On-Chip Wires Through Capacitive Coupling," patent pending, filed 9/29/05.
[7] B. Nikolic, V.G. Oklobdzija, V. Stojanovic, et al., "Improved Sense-Amplifier-Based Flip-Flop," *IEEE J. Solid State Circuits*, vo. 35, no. 6, pp. 876-884, Jun., 2000.
[8] R. Hopkins, et al. "Equalization in Capacitively-Coupled Links," Patent Pending, filed May 2006.
[9] R. Ho, B. Amrutur, K. Mai, et al., "Applications of On-Chip Samplers for Test and Measurement of Integrated Circuits," *IEEE Symp. VLSI Circuits*, pp. 138-139, Jun., 1998

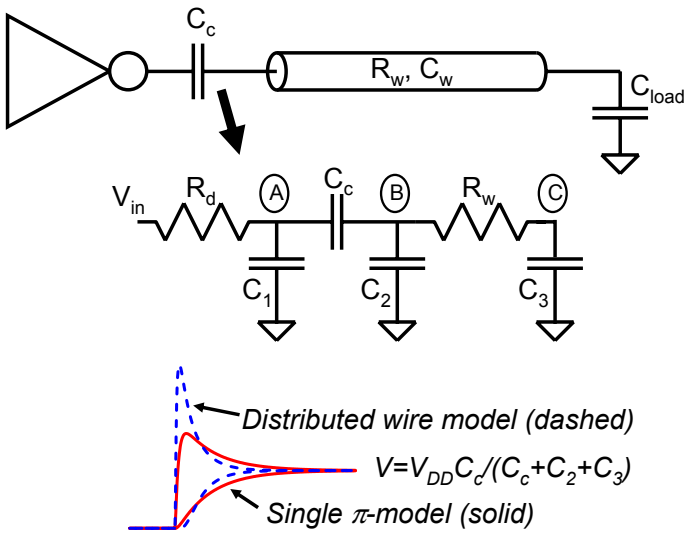

Figure 22.8.1: Simple model of a capacitively driven wire.

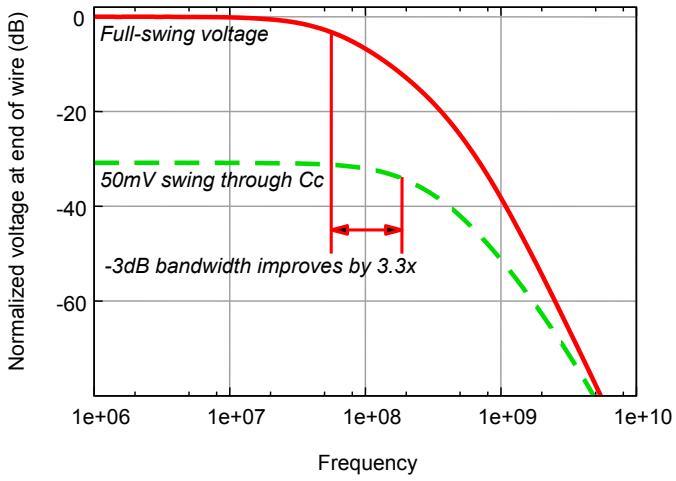

Figure 22.8.2: Capacitor extends bandwidth on a 14mm minimum-width wire.

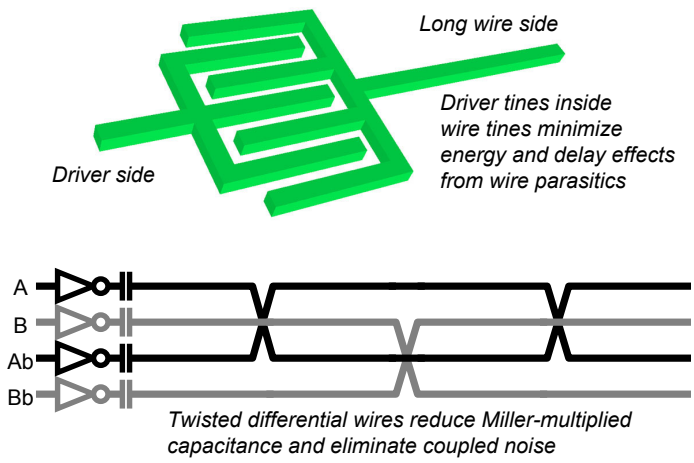

Figure 22.8.3: Capacitor construction and wire twisting.

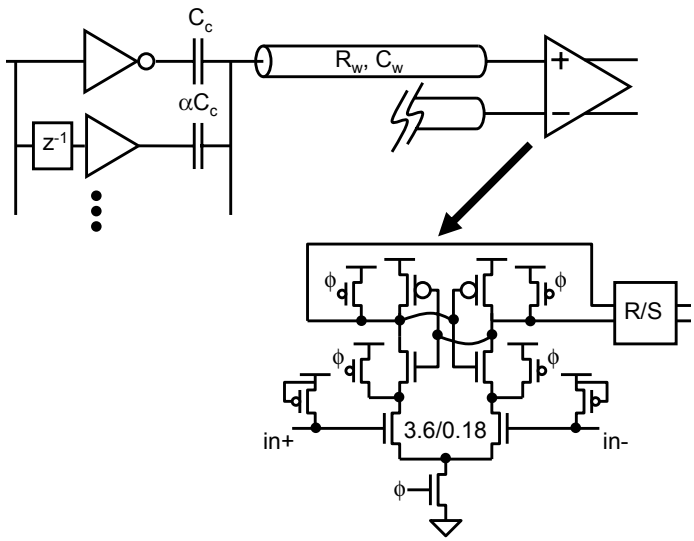

Figure 22.8.4: Creating a simple multi-tap FIR filter and circuit details.

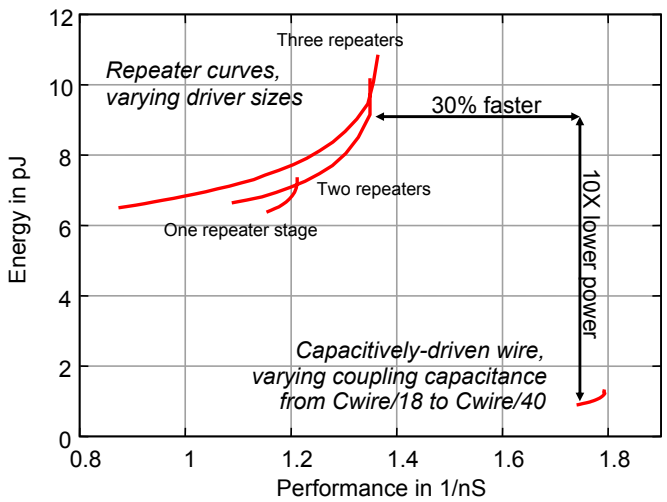

Figure 22.8.5: Simulated comparison with repeaters for a 10mm long bus.

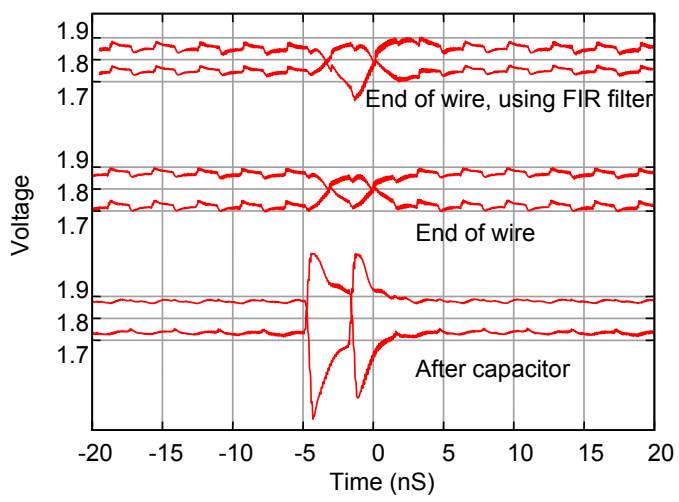

Figure 22.8.6: Measured pre-emphasis on a slow minimum-width 14mm bus.

Continued on Page 612

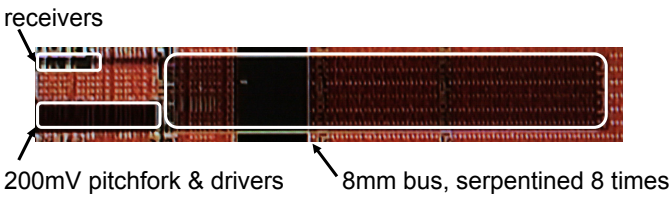

Figure 22.8.7: Die micrograph of a representative low-swing bus.

